1. Общая схема постановки задачи

Важно: рассматриваем только задачу классификации для простоты.

1) Формально расписываем что у нас есть (непараллельные корпуса) и какая модель у нас является переводчиком (decoder от RNN)

2) Описываем, какой функционал хотелось бы минимизировать (можно описать в терминах классификации каждого слова при переводе)

3) Объясняем, что поскольку параллельных данных нет, то предлагается свести задачу к нахождению общего пространства для двух языков. Здесь нужно формально выписать свойство, которым описывается это общее пространство.

2. Общие обозначения

- Обучающая выборка на первом языке: $\mathfrak{D}^{\text{src}} = [\mathbf{s}_1^{\text{src}}, \dots, \mathbf{s}_{m_{\text{src}}}^{\text{src}}]$.
- Обучающая выборка на втором языке: $\mathfrak{D}^{\text{tgt}} = [\mathbf{s}_1^{\text{tgt}}, \dots, \mathbf{s}_{m_{\text{tgt}}}^{\text{tgt}}]$.
- Максимальная длина предложения: l.
- Мощность словаря на первом языке: V^{src}.
- Мощность словаря на втором языке: V^{tgt}.
- Предложение на первом языке: $\mathbf{s}^{\text{src}} = [x_1, \dots, x_l], \quad x_i \in \{1, \dots, V^{\text{src}}\}$.
- Предложение на втором языке: $\mathbf{s}^{\text{tgt}} = [y_1, \dots, y_l], \quad y_i \in \{1, \dots, V^{\text{tgt}}\}$.
- Энкодер: $\mathbf{f}$. (Если потребуется — ставьте индекс src или tgt около модели).
- Декодер: $\mathbf{g}$. (Если потребуется — ставьте индекс src или tgt около модели).
- Скрытое представление: $\mathbf{h}$.
- Валидационная выборка: $\mathfrak{D}^{valid} = \{(\mathbf{s}_1^{\text{src}}, \mathbf{s}_1^{\text{tgt}}), \dots, (\mathbf{s}_{m_{\text{valid}}}, \mathbf{s}_{\text{valid}})\}$.
- Параметры Seq2Seq-модели: $\mathbf{w}$.